On Slepian–Wolf Theorem with Interaction

Alexander Kozachinskiy*

*Moscow State University, Faculty of Mechanics and Mathematics
kozlach@mail.ru

September 9, 2015

Abstract

In this paper we study interactive "one-shot" analogues of the classical Slepian-Wolf theorem. Alice receives a value of a random variable X, Bob receives a value of another random variable Y that is jointly distributed with X. Alice's goal is to transmit X to Bob (with some error probability ε). Instead of one-way transmission, which is studied in the classical coding theory, we allow them to interact. They may also use shared randomness.

We show, that Alice can transmit X to Bob in expected $H(X|Y) + 2\sqrt{H(X|Y)} + O(\log_2\left(\frac{1}{\varepsilon}\right))$ number of bits. Moreover, we show that every one-round protocol π with information complexity I can be compressed to the (many-round) protocol with expected communication about $I + 2\sqrt{I}$ bits. This improves a result by Braverman and Rao [3], where they had $5\sqrt{I}$. Further, we show how to solve this problem (transmitting X) using $3H(X|Y) + O(\log_2\left(\frac{1}{\varepsilon}\right))$ bits and 4 rounds on average. This improves a result of [4], where they had $4H(X|Y) + O(\log 1/\varepsilon)$ bits and 10 rounds on average.

In the end of the paper we discuss how many bits Alice and Bob may need to communicate on average besides $H(X|Y)$. The main question is whether the upper bounds mentioned above are tight. We provide an example of (X, Y), such that transmission of X from Alice to Bob with error probability ε requires $H(X|Y) + \Omega\left(\log_2\left(\frac{1}{\varepsilon}\right)\right)$ bits on average.

1 Introduction

Assume that Alice receives a value of a random variable X and she wants to transmit that value to Bob. It is well-known ([8]) that Alice can do it using one message over the binary alphabet of expected length less than $H(X)+1$. Assume now that there are n independent random variables $X_1, \ldots, X_n$ distributed as X, and Alice wants to transmit all $X_1, \ldots, X_n$ to Bob. Another classical result from [8] states, that Alice can do it using one message of *fixed* length, namely $\approx nH(X)$, with a small probability of error.

One of the possible ways to generalize this problem is to provide Bob with a value of another random variable Y which is jointly distributed with X. That is, to let Bob know some partial information about X for free. This problem is the subject of the classical Slepian-Wolf Theorem [9] which asserts that if there are n independent pairs $(X_1, Y_1), \ldots, (X_n, Y_n)$, each pair distributed exactly as (X, Y), then Alice can transmit all $X_1, \ldots, X_n$ to Bob, who knows $Y_1, \ldots, Y_n$, using one message of fixed length, namely $\approx nH(X|Y)$, with a small probability of error. However, it turns out that a one-shot analogue of this theorem is impossible, if only one-way communication is allowed.

The situation is quite different, if we allow Alice and Bob to *interact*, that is, to send messages in both directions. In [7] Orlitsky studied this problem for the average-case communication when no error is allowed. He showed that if pair (X, Y) is uniformly distributed on it's support, then Alice may transmit X to Bob using at most

$$H(X|Y) + 3\log_2(H(X|Y) + 1) + 17$$

bits on average and 4 rounds. For the pairs (X, Y) whose support is a Cartesian product Orlitsky showed that error-less transmission of X from Alice to Bob requires $H(X)$ bits on average.

From a result of Braverman and Rao ([3]), it follows that for arbitrary (X, Y) it is sufficient to communicate at most

$$H(X|Y) + 5\sqrt{H(X|Y)} + O\left(\log_2\left(\frac{1}{\varepsilon}\right)\right)$$

bits on average (here ε stands for the error probability).

We improve this result, showing that Alice may transmit X to Bob with error probability at most ε (for each pair of inputs) using at most

$$H(X|Y) + 2\sqrt{H(X|Y)} + O\left(\log_2\left(\frac{1}{\varepsilon}\right)\right)$$

bits on average and $O(\sqrt{H(X|Y)})$ rounds. Our protocol is inspired by protocol from [1]. The idea of the protocol is essentially the same, we only apply some technical trick to reduce communication.

Actually, in [3] a more general result was established. It was shown there that every one-round protocol π with information complexity I can be compressed to the (many-round) protocol with expected length at most

$$\approx I + 5\sqrt{I}. \tag{1}$$

Using the result from [2], we improve 1. Namely, we show that every one-round protocol π with information complexity I can be compressed to the (many-round) protocol with expected communication length at most

$$\approx I + 2\sqrt{I}.$$

In [4], it is established a one-shot interactive analogue of the Slepian-Wolf theorem for the bounded-round communication. They showed that Alice may transmit X to Bob using at most $O(H(X|Y)+1)$ bits and $O(1)$ rounds on average. More specifically, their protocol transmits at most $4H(X|Y)+\log_2(1/\varepsilon)+O(1)$ bits on average in 10 rounds on average. In this paper, we provide another proof of this result, which seems to be easier. More specifically, we show that it is sufficient to communicate at most

$$3H(X|Y)+\log_2\left(\frac{1}{\varepsilon}\right)+O(1)$$

bits on average in at most 4 rounds on average.

From the proof of our upper bound it follows that there exists a *deterministic protocol* which transmits X from Alice to Bob using the same number of bits on average (namely $H(X|Y)+2\sqrt{H(X|Y)}+O\left(\log_2\left(\frac{1}{\varepsilon}\right)\right)$) and which guaranties that for at most ε-fraction of inputs (with respect to the distribution of (X,Y)) the transmission is incorrect. Are there random variables X,Y for which the corresponding upper bound is tight? We make a step towards answering this question: we provide an example of random variables X,Y such that every deterministic protocol which transmits X from Alice to Bob with error probability ε must communicate at least $H(X|Y)+\Omega\left(\log_2\left(\frac{1}{\varepsilon}\right)\right)$ bits on average.

In the Appendix we provide an example of (X,Y) for which it seems plausible that the upper bound $H(X|Y)+O(\sqrt{H(X|Y)})$ is tight.

2 Definitions

We will denote the set of the first n naturals $\{1,2,\ldots,n\}$ by $[n]$.

2.1 Information Theory

Let X, Y be two joint distributed random variables, taking values in the finite sets, respectively, $\mathcal{X}$ and $\mathcal{Y}$.

Definition 2.1. *Shannon Entropy of X is defined by the formula*

$$H(X)=\sum_{x\in\mathcal{X}}\Pr[X=x]\log_2\left(\frac{1}{\Pr[X=x]}\right).$$

Definition 2.2. *Conditional Shannon entropy of X with respect to Y is defined by the formula:*

$$H(X|Y)=\sum_{y\in\mathcal{Y}}H(X|Y=y)\Pr[Y=y],$$

where $X|Y=y$ denotes a distribution of X, conditioned on the event $\{Y=y\}$.

If X is uniformly distributed in $\mathcal{X}$ then obviously $H(X)=\log_2(|\mathcal{X}|)$. We will also use the fact that the formula for conditional entropy may be re-written

as

$$H(X|Y) = \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \Pr[X=x, Y=y] \log_2 \left(\frac{1}{\Pr[X=x|Y=y]} \right).$$

Generalization of the Shannon entropy is Renyi entropy.

Definition 2.3. *Renyi entropy of X is defined by the formula*

$$H_2(X) = -\log_2 \left(\sum_{x\in\mathcal{X}} \Pr[X=x]^2 \right).$$

Concavity of log implies that $H(X) \geq H_2(X)$.

The mutual information of two random variables X and Y, conditioned on another random variable Z, can be defined as:

$$I(X:Y|Z) = H(X|Z) - H(X|Y,Z).$$

For the further introduction in information theory see, for example [11].

2.2 Communication Protocols

Assume that we are given jointly distributed random variables X and Y, taking values in finite sets $\mathcal{X}$ and $\mathcal{Y}$. Let R, R_A, R_B be a random variables, taking values in finite sets $\mathcal{R}$, $\mathcal{R}_\mathcal{A}$ and $\mathcal{R}_\mathcal{B}$, such that $(X,Y), R, R_A, R_B$ are mutually independent.

Definition 2.4. *A randomized communication protocol is a rooted binary tree, in which each non-leaf vertex is associated either with Alice or with Bob. For each non-leaf vertex v associated with Alice there is a function $f_v : \mathcal{X}\times\mathcal{R}\times\mathcal{R}_\mathcal{A} \to \{0,1\}$ and for each non-leaf vertex u associated with Bob there is a function $g_u : \mathcal{Y}\times\mathcal{R}\times\mathcal{R}_\mathcal{B} \to \{0,1\}$. For each non-leaf vertex one of an out-going edges is labeled by 0 and other is labeled by 1. Finally, for each leaf l there is a function $\phi_l : \mathcal{Y}\times\mathcal{R}\times\mathcal{R}_\mathcal{B} \to \mathcal{O}$, where $\mathcal{O}$ denotes the set of all possible Bob's outputs.*

A computation according to a protocol runs as follows. Alice is given $x \in \mathcal{X}$, Bob is given $y \in \mathcal{Y}$. Assume that the random variables R takes a value r, R_A takes a value r_a and R_B takes a value r_b. Alice and Bob start at the root of the tree. If they are in the non-leaf vertex v associated with Alice, then Alice sends $f_v(x,r,r_a)$ to Bob and they go by the edge labeled by $f_v(x,r,r_a)$. If they are in a non-leaf vertex associated with Bob then Bob sends $g_v(y,r,r_b)$ to Alice and they go by the edge labeled by $g_v(y,r,r_b)$. When they reach a leaf l Bob outputs the result $\phi_l(y,r,r_b)$.

A protocol is called *public-coin* if f_v, g_u and ϕ_l do not depend on the values of R_A, R_B.

A protocol is called *deterministic* if f_v, g_u and ϕ_l do not depend on the values of R, R_A, R_B.

We distinguish between average-case communication complexity and the worst-case communication complexity.

Definition 2.5. *The (worst-case) communication complexity of a protocol π, denoted by $CC(\pi)$, is defined as the depth of the corresponding binary tree.*

We say that protocol π communicates d bits on average (or expected length of the protocol is equal to d), if the expected depth of the leaf that Alice and Bob reach during the execution of the protocol π is equal to d, where the expectation is taken over X, Y, R, R_A, R_B.

If the Alice's goal is to transmit X to Bob, then in the end of the communication Bob should output some element of $\mathcal{X}$ (that is, $\mathcal{O} = \mathcal{X}$). We say that protocol transmits X from Alice to Bob with error probability ε if

$$\Pr[X = \phi_L(Y, R, R_B)] \geq 1 - \varepsilon,$$

where L denotes the leaf that Alice and Bob reach in the protocol tree.

For the worst-case communication it is sufficient to consider only deterministic protocols. Indeed, assume that we are given a randomized protocol solving our problem with error probability ε. Fix the value of R for which error probability is minimal. In this way we obtain a protocol with the same worst-case communication complexity and error probability.

For the further introduction in Communication Complexity see [5]

3 Near-optimal one-shot Slepian-Wolf theorem

Consider the following auxiliary problem. Let A be a finite set. Assume that Alice receives an arbitrary $a \in A$ and Bob receives and arbitrary probability distribution μ on A. Alice wants to communicate a to Bob in about $\log(1/\mu(a))$ bits with small probability of error.

Lemma 3.1. *Let ε be a positive real and h a positive integer. There exists a public coin randomized communication protocol such that for all a in the support of μ the following hold:*

- *in the end of the communication Bob outputs $b \in A$ which is equal to a with probability at least $1 - \varepsilon$;*
- *the protocol communicates at most*

$$\log_2\left(\frac{1}{\mu(a)}\right) + \frac{\log_2\left(\frac{1}{\mu(a)}\right)}{h} + h + \log_2\left(\frac{1}{\varepsilon}\right) + O(1)$$

bits, regardless of the randomness.

Proof. Alice and Bob interpret each portion of $|A|$ consecutive bits from the public randomness source as a table of a random function $h : A \to \{0, 1\}$. That is, we will think that they have access to a large enough family of mutually independent random functions of the type $A \to \{0, 1\}$. Those functions will be called *hash functions* and their values *hash values* below.

The first set $k = \left\lceil \log_2 \left(\frac{1}{\varepsilon}\right) \right\rceil + 1$. Then Bob sets:

$$S_i = \left\{ x \in A \,|\, \mu(x) \in (2^{-i-1}, 2^{-i}] \right\}.$$

Then Alice and Bob work in stages numbered $0, 1, \ldots$.
On *Stage 0:*

1. Alice sends k hash values of a to Bob.
2. Bob computes set S_0', which consists of all elements from S_0 that have the same hash values as sent by Alice (actually S_0 has at most one element).
3. If $S_0' \neq \varnothing$, then Bob sends 1 to Alice, outputs any element of S_0' and they terminate. Otherwise Bob sends 0 to Alice and they proceed of Stage 1.

On *Stage t*:

1. Alice sends h new hash values of a to Bob so that the total number of hash values of a available to Bob be $k + ht$.
2. For each $i \in \{h(t-1)+1, \ldots, ht\}$ Bob computes set S_i', which consists of all elements from S_i, which agree with all Alice's hash values.
3. If there exists $i \in \{h(t-1)+1, \ldots, ht\}$ such that $S_i' \neq \varnothing$, then Bob sends 1 to Alice, outputs any element of S_i' and they terminate. Otherwise Bob sends 0 to Alice and they proceed to Stage $t+1$.

Let us at first show that the protocol terminates for all a in the support of μ. Assume that Alice has a and Bob has μ. Let $i = \left\lfloor \log_2 \left(\frac{1}{\mu(a)}\right) \right\rfloor$ so that $a \in S_i$. The protocol terminates on Stage t where

$$h(t-1) + 1 \leq i \leq ht$$

or earlier. Indeed all hash values of a available to Bob on Stage t coincide with hash values of some element of S_i (for instance, with those of a).

Thus Alice sends at most $k + ht$ bits to Bob and Bob sends at most $1 + t$ bits to Alice. Therefore total communication is bounded by

$$\begin{aligned} k + ht + 1 + t &= k + h(t-1) + h + 2 + (t-1) \\ &\leq k + i - 1 + h + 2 + \frac{i-1}{h} \\ &\leq k + \log_2 \left(\frac{1}{\mu(a)}\right) + \frac{\log_2 \left(\frac{1}{\mu(a)}\right)}{h} + h + O(1). \end{aligned}$$

Since $k = \left\lceil \log_2 \left(\frac{1}{\varepsilon}\right) \right\rceil + 1$, the required bound follows.

Now we bound the error probability. An error may occurs, if for some t a set S_i considered on Stage t has an element $b \neq a$ which agrees with hash values sent from Alice. At that time Bob has already $k + ht \geq k + i$ hash values. The

probability that $k+i$ hash values of b coincide with those of a is 2^{-k-i}. Hence by union bound error probability does not exceed

$$\sum_{i=0}^{\infty} |S_i| 2^{-k-i} = 2^{-k+1} \sum_{i=0}^{\infty} |S_i| 2^{-i-1} < 2^{-k+1} \sum_{i=0}^{\infty} \sum_{x \in S_i} \mu(x)$$
$$= 2^{-k+1} \sum_{x \in A} \mu(x) = 2^{-k+1} = 2^{-\lceil \log_2 (\frac{1}{\varepsilon}) \rceil} \leq \varepsilon.$$

□

Theorem 3.1. *Let X, Y be jointly distributed random variables that take values in the finite sets $\mathcal{X}$ and $\mathcal{Y}$. Then for every positive ε there exists a public-coin protocol with the following properties.*

- *For every pair (x, y) from the support of (X, Y) with probability at least $1 - \varepsilon$ Bob outputs x;*
- *The expected length of communication is at most*

$$H(X|Y) + 2\sqrt{H(X|Y)} + \log_2 \left(\frac{1}{\varepsilon}\right) + O(1).$$

Proof. On input x, y, Alice and Bob run protocol of Lemma 3.1 with $A = \mathcal{X}$, $h = \left\lceil \sqrt{H(X|Y)} \right\rceil$, $a = x$ and μ equal to the distribution of X, conditioned on the event $Y = y$. Notice that Alice knows a and Bob knows μ.

Let us show that both requirements are fulfilled for this protocol. The first requirement immediately follows from the first property of the protocol of Lemma 3.1.

From the second property of the protocol of Lemma 3.1 it follows that for input pair x, y out protocol communicates at most:

$$\log_2 \left(\frac{1}{\Pr[X = x | Y = y]}\right) + \frac{\log_2 \left(\frac{1}{\Pr[X=x|Y=y]}\right)}{\left\lceil \sqrt{H(X|Y)} \right\rceil} + \left\lceil \sqrt{H(X|Y)} \right\rceil + \log_2 \left(\frac{1}{\varepsilon}\right) + O(1)$$

bits. Recalling that

$$H(X|Y) = \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \Pr[X = x, Y = y] \log_2 \left(\frac{1}{\Pr[X = x | Y = y]}\right)$$

we see on average the communication is as short as required. □

Remark. One may wonder whether there exists a private-coin communication protocol with the same properties as the protocol of Theorem 3.1. Newman's theorem ([6]) states that every public-coin protocol can be transformed into a private-coin protocol at the expense of increasing the error probability by δ and the worst case communication by $O(\log \log |\mathcal{X} \times \mathcal{Y}| + \log 1/\delta)$ (for any

positive δ). Lemma 3.1 provides an upper bound for the error probability and communication of our protocol for each pair of inputs. Repeating the arguments from the proof of Newman's theorem, we are able to transform the public-coin protocol of Lemma 3.1 into a private-coin one with the same trade off between the increase of error probability and the increase of communication length. It follows that for our problem there exists a private-coin communication protocol which errs with probability at most ε and communicates on average as many bits as the public-coin protocol from Theorem 3.1 plus extra $O(\log\log|\mathcal{X}\times\mathcal{Y}|)$ bits.

4 One-shot Slepian-Wolf theorem with a constant number of rounds on average

In this section, we modify the construction from the previous section to reduce the average number of rounds to a constant.

Theorem 4.1. *Let X, Y be jointly distributed random variables that take values in the finite sets $\mathcal{X}$ and $\mathcal{Y}$. Then for every positive ε there exists a public-coin protocol with the following properties:*

- *For every pair (x,y) from the support of (X,Y) with probability at least $1-\varepsilon$ Bob outputs x;*
- *The expected length of the protocol does not exceed*

$$3H(X|Y)+\log_2\left(\frac{1}{\varepsilon}\right)+O(1).$$

- *The expected number of rounds in protocol is at most 4.*

(Compared to Theorem 3.1, the number of rounds has decreased and the communication length has increased.)

Proof. We will use the following notation:

$$l=\lceil H(X|Y)\rceil,\qquad k=\left\lceil\log_2\left(\frac{1}{\varepsilon}\right)\right\rceil+1,$$

$$\mu(x,y)=\Pr[X=x,Y=y],\qquad \mu(x|y)=\Pr[X=x|Y=y].$$

Alice and Bob apply the following modification of the protocol of Lemma 3.1. Recall that that protocol works in stages. On Stage 0 Alice sends to Bob k random hash bits and on each subsequent stage Alice sends to Bob extra h random hash bits. On each stage Bob looks for an element in all sets

$$S_i=\{x'\mid\mu(x'|y)\in(2^{-i-1},2^{-i}]\}.$$

such that i is at least k less than the total number of hash bits he has so far. This guarantees that the error probability is at least ε for all input pairs.

Now Alice sends $k+l$ hash bits on Stage 0 and $l2^t$ new hash bits on Stage $t>0$. This is the main difference between the new protocol and the protocol of Theorem 3.1. In order to keep the error probability at most ε, on Stage t Bob looks for an element in S_i with the same hash values as sent by Alice for $i \le l+2l+\ldots+2^t l$. If there is such an element, then Bob outputs any such element (and sends 1 to Alice). Otherwise he sends 0 and they proceed to the next stage.

As earlier, by union bound the error probability does not exceed

$$\sum_{i=0}^{\infty} |S_i| 2^{-k-i} = 2^{-k+1} \sum_{i=0}^{\infty} |S_i| 2^{-i-1} \le 2^{-k+1} \sum_{i=0}^{\infty} \sum_{x' \in S_i} \mu(x'|y)$$
$$= 2^{-k+1} \sum_{x' \in \mathcal{X}} \mu(x'|y) \le 2^{-k+1} = 2^{-\lceil \log_2 \left(\frac{1}{\varepsilon}\right) \rceil} \le \varepsilon.$$

Now we will estimate the communication length on each input pair (x,y) of positive probability. Bob sends one bit in each round. As we will see the average number of rounds is at most 4, thus we may forget about the communication from Bob and concentrate on communication from Alice.

Set $j = j(x,y) = \left\lfloor \log_2 \left(\frac{1}{\mu(x|y)} \right) \right\rfloor$. Notice that $x \in S_j$. Consider t such that

$$l + 2l + \ldots + 2^{t-1} l < j \le l + 2l + \ldots + 2^t l. \tag{2}$$

By the construction of the protocol the communication length for input x, y is at most

$$\begin{aligned} & k + l + 2l + \ldots + 2^t l \\ & = k + l + 2(l + 2l + \ldots + 2^{t-1} l) \\ & < k + l + 2j. \end{aligned}$$

Hence the expected length of communication from Alice to Bob is at most

$$\begin{aligned} & \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \mu(x,y)(k + l + 2j(x,y)) \\ & \le k + l + 2 \sum_{(x,y) \in \mathcal{X} \times \mathcal{Y}} \mu(x,y) \log_2 \left(\frac{1}{\mu(x|y)} \right) \\ & = k + l + 2H(X|Y) = 3H(X|Y) + k + O(1). \end{aligned}$$

Let us bound the expected number of rounds in our protocol. Let $R(x,y)$ stand for the number for inputs $X = x$, $Y = y$. Then $R(x,y)$ is at most $2t+2$, where t is defined by (2). By (2) we have

$$(2^t - 1) l < j \le \log_2 \left(\frac{1}{\mu(x|y)} \right)$$

and hence:

$$t \le \log_2 \left(1 + \frac{\log_2 \left(\frac{1}{\mu(x|y)} \right)}{l} \right).$$

Thus:

$$R(x,y) \leq 2 + 2\log_2\left(1 + \frac{\log_2\left(\frac{1}{\mu(x|y)}\right)}{l}\right).$$

By concavity of the logarithmic function the average number of rounds does not exceed:

$$\begin{aligned}\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \mu(x,y)R(x,y) &\leq \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \mu(x,y)\left(2 + 2\log_2\left(1 + \frac{\log_2\left(\frac{1}{\mu(x|y)}\right)}{l}\right)\right)\\ &\leq 2 + 2\log_2\left(\sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \mu(x,y)\left(1 + \frac{\log_2\left(\frac{1}{\mu(x|y)}\right)}{l}\right)\right)\\ &\leq 2 + 2\log_2(2) = 4.\end{aligned}$$

□

5 One-round Compression

Information complexity of the protocol π with inputs (X,Y) is defined as

$$\begin{aligned}IC_\mu(\pi) &= I(X:\Pi|Y,R) + I(Y:\Pi|X,R)\\ &= I(X:\Pi|Y,R,R_B) + I(Y:\Pi|X,R,R_A)\\ &= I(X:\Pi,R,R_B|Y) + I(Y:\Pi,R,R_A|X),\end{aligned}$$

where R, R_A, R_B denote (shared, Alice's and Bob's) randomness, μ stands for the distribution of (X,Y) and Π stands for the concatenation of all bits sent in π (Π is called a *transcript*). The first term is equal to the information which Bob learns about Alice's input and the second term is equal to the information which Alice learns about Bob's input. Information complexity is an important concept in the Communication Complexity. For example, information complexity plays the crucial role in the Direct-Sum problem ([10]).

We will consider the special case when π is *one-round*. In this case Alice sends one message Π to Bob, then Bob outputs the result (based on his input, his randomness, and Alice's message) and the protocol terminates. Since Alice learns nothing, information complexity can be re-written as

$$I = IC_\mu(\pi) = I(X:\Pi|Y,R).$$

Our goal is to simulate a given one-round protocol π with another protocol τ which has the same input space (X,Y) and whose expected communication complexity is close to I. The new protocol τ may be many-round. The quality of simulation will be measured by the statistical distance. Statistical distance between random variables A and B, both taking values in the set V, equals

$$\delta(A,B) = \max_{U\subset V} \left|\Pr[A\in U] - \Pr[B\in U]\right|.$$

One of the main results of [3] is the following theorem.

Theorem 5.1. *For every one-round protocol π and for every probability distribution μ there is a public-coin protocol τ with expected length (with respect to μ and the randomness of τ) at most $I + 5\sqrt{I} + O\left(\log_2 \frac{1}{\varepsilon}\right)$ such that for each pair of inputs (x, y) after termination of τ Bob outputs a random variable Π' with $\delta\left((\Pi|X = x, Y = y), (\Pi'|X = x, Y = y)\right) \leq \varepsilon$.*

We will show that theorem 3.1 implies that we can replace $5\sqrt{I}$ by about $2\sqrt{I}$ in this theorem. We want transmit Alice's message Π to Bob (who knows Y and his randomness R) in many rounds so that the expected communication length is small. By theorem 3.1 this task can be solved with error ε in expected communication

$$H(\Pi|Y, R) + 2\sqrt{H(\Pi|Y, R)} + O\left(\log_2 \frac{1}{\varepsilon}\right). \tag{3}$$

Assume first that the original protocol π uses only public randomness. Then

$$I = I(X : \Pi|Y, R) = H(\Pi|Y, R) - H(\Pi|X, Y, R) = H(\Pi|Y, R).$$

Indeed, $H(\Pi|X, Y, R) = 0$, since Π is defined by X, R. Thus (3) becomes

$$I + 2\sqrt{I} + O\left(\log_2 \frac{1}{\varepsilon}\right)$$

and we are done.

Fortunately, by the following theorem from [2] we can remove private coins from the protocol with only a slight increase in information complexity.

Theorem 5.2. *there is a one-round public-coin protocol π' with information complexity $IC_\mu(\pi) \leq I + \log_2(I + O(1))$ such that for each pairs of inputs (x, y) Bob outputs Π' for which $\Pi'|X = x, Y = y$ and $\Pi|X = x, Y = y$ are identically distributed.*

Combining this theorem with our main result (theorem 3.1), we obtain the following theorem.

Theorem 5.3. *there is a public-coin protocol τ with expected length (with respect to μ and the randomness of τ) at most*

$$I + \log_2(I + O(1)) + 2\sqrt{I + \log_2(I + O(1))} + O\left(\log_2 \frac{1}{\varepsilon}\right)$$

such that for each pair of inputs (x, y) in Bob outputs Π' $\delta\left((\Pi|X = x, Y = y), (\Pi'|X = x, Y = y)\right) \leq \varepsilon$

6 A Lower Bounds for the Average-Case Communication

Let (X, Y) be a pair of jointly distributed random variables. Assume that π is a deterministic protocol to transmit X from Alice to Bob who knows Y. Let $\pi(X, Y)$ stand for the result output by the protocol π for input pair (X, Y). We assume that for at least $1 - \varepsilon$ input pairs this result is correct:

$$\Pr[\pi(X, Y) \neq X)] \leq \varepsilon.$$

It is not hard to see that in this case the expected communication length cannot be much less than $H(X|Y)$ bits on average. Moreover, this applies for communication from Alice to Bob only.

Proposition 6.1. *For every deterministic protocol as above the expected communication from Alice to Bob is at least* $H(X|Y) - \varepsilon \log_2 |\mathcal{X}| - 1$.

Proof. Indeed, let Π_A denote the concatenation of all bits sent by Alice. If Bob's input is fixed, then the set of all possible values of Π_A forms a prefix-free code. Hence

$$\mathsf{E}\left[|\Pi_A| \mid Y = y\right] \geq H(\Pi_A|Y = y)$$

and therefore

$$\mathsf{E}|\Pi_A| = \mathsf{E}_{y\sim Y}\mathsf{E}\left[|\Pi_A| \mid Y = y\right] \geq \mathsf{E}_{y\sim Y}H(\Pi_A|Y = y) = H(\Pi_A|Y).$$

Consider $I(X : \Pi_A|Y)$. By definition $I(X : \Pi_A|Y) \leq H(\Pi_A|Y)$. On the other hand we have

$$I(X : \Pi_A|Y) = H(X|Y) - H(X|Y, \Pi_A).$$

Notice that $\pi(X, Y)$ is a function of Y and π_A (Bob's guess is based on Y and on bits received from Alice) and hence $H(X|Y, \Pi_A) \leq H(X|\pi(X, Y))$. Since $\Pr[\pi(X, Y) \neq X] \leq \varepsilon$, from Fano inequality it follows that

$$H(X|\pi(X, Y)) \leq 1 + \varepsilon \log_2 |\mathcal{X}|.$$

Therefore $\mathsf{E}|\Pi_A| \geq H(X|Y) - \varepsilon \log_2 |\mathcal{X}| - 1$. □

There are random variables for which this lower bound is tight. For instance, let Y be empty and let X take the value $x \in \{0, 1\}^n$ with probability $\varepsilon/2^n$ (for all such x) and let $X =$ (the empty string) with the remaining probability $1 - \varepsilon$. Then the trivial protocol with no communication solves the job with error probability ε and $H(X|Y) \approx \varepsilon \log_2 |\mathcal{X}|$.

In this section we consider the following question: are there a random variables (X, Y), for which for every deterministic communication protocol the expected communication is significantly larger than $H(X|Y)$, say close to the upper bound $H(X|Y) + 2\sqrt{H(X|Y)} + \log_2\left(\frac{1}{\varepsilon}\right)$ of Theorem 3.1? Notice that from the proof of the theorem 3.1 it follows that there exists a *deterministic protocol* which transmits X from Alice to Bob using $H(X|Y)+2\sqrt{H(X|Y)}+O\left(\log_2\left(\frac{1}{\varepsilon}\right)\right)$

bits on average and which guaranties that for at most ε-fraction of inputs (with respect to the distribution of (X, Y)) the transmission is incorrect. Indeed, for any choice of randomness the communication on each pair of inputs is bounded by lemma 3.1. Thus we may fix random bits so that the error probability is at most ε.

Orlitsky showed that if no error is allowed and the support of (X, Y) is a Cartesian product, then every deterministic protocol must communicate $H(X)$ bits on average.

Lemma 6.1. *Let (X, Y) be a pair of jointly distributed random variables whose support is a Cartesian product. Assume that π is a deterministic protocol, which transmits X from Alice to Bob who knows Y and*

$$\Pr[\pi(X, Y) \neq X)] = 0.$$

Then the expected length of π is at least $H(X)$.

For the sake of completeness we provide a proof of this result in the Appendix. The main result of this section states that there are random variables (X, Y) such that transmission of X from Alice to Bob with error probability ε requires $H(X|Y) + \Omega\left(\log_2\left(\frac{1}{\varepsilon}\right)\right)$ bits on average.

The random variables X, Y are specified by two parameters, $\delta \in (0, 1/2)$ and $n \in \mathbb{N}$. Both random variables take values in $\{0, 1, \ldots, n\}$ and are distributed as follows: Y is distributed uniformly in $\{0, 1, \ldots, n\}$ and $X = Y$ with probability $1 - \delta$ and X is uniformly distributed in $\{0, 1, \ldots, n\} \setminus \{X\}$ with the remaining probability δ. That is,

$$\Pr[X = i, Y = j] = \frac{(1 - \delta)\delta_{ij} + \frac{\delta}{n}(1 - \delta_{ij})}{n + 1},$$

where δ_{ij} stands for the Kronecker's delta. Notice that X is uniformly distributed on $\{0, 1, \ldots, n\}$ as well. A straightforward calculation reveals that

$$\Pr[X = i|Y = j] = \frac{\Pr[X = i, Y = j]}{\Pr[Y = j]} = (1 - \delta - \frac{\delta}{n})\delta_{ij} + \frac{\delta}{n}$$

and

$$H(X|Y) = (1 - \delta)\log_2\left(\frac{1}{1 - \delta}\right) + \delta\log_2\left(\frac{n}{\delta}\right) = \delta\log_2 n + O(1).$$

We will think of δ as a constant, say $1/4$. For one-way protocol we are able to show that communication length must be close to $\log n$, which is about $1/\delta$ times larger than $H(X|Y)$:

Proposition 6.2. *Assume that π is a one-way deterministic protocol, which transmits X from Alice to Bob who knows Y and*

$$\Pr[\pi(X, Y) \neq X)] \leq \varepsilon.$$

Then the expected length of π is at least $\left(1 - \frac{\varepsilon}{\delta}\right)\log_2(n + 1) - 2$.

Proof. Let S be the number of leafs in π. For each $j \in \{0, 1, \ldots, n\}$

$$\#\left\{i \in \{0, 1, \ldots, n\} \mid \pi(i, j) = i\right\} \leq S.$$

Hence the error probability ε is at least $(n + 1 - S)\frac{\delta}{n}$. This implies that

$$S \geq n\left(1 - \frac{\varepsilon}{\delta}\right) + 1 \geq (n + 1)\left(1 - \frac{\varepsilon}{\delta}\right).$$

Let $\Pi(X)$ denote the leaf Alice and Bob reach in π (since the protocol is one-way, the leaf depends only on X). The expected length of $\Pi(X)$ is at least $H(\Pi)$. Let $l_1, l_2, \ldots, l_S$ be the list of all leaves in the support of the random variable $\Pi(X)$. As X is distributed uniformly, we have

$$\Pr[\Pi = l_i] \geq \frac{1}{n + 1}$$

for all i. The statement follows from

Lemma 6.2. *Assume that $p_1, \ldots, p_k, q_1, \ldots, q_k \in (0, 1)$ satisfy*

$$\sum_{i=1}^{k} p_i = 1,$$

$$\forall i \in \{1, \ldots, k\} \qquad p_i \geq q_i.$$

Then

$$\sum_{i=1}^{k} p_i \log_2 \frac{1}{p_i} \geq \sum_{i=1}^{k} q_i \log_2 \frac{1}{q_i} - 2.$$

The proof of this technical lemma is deferred to the Appendix. The lemma implies that

$$\begin{aligned} H(\Pi) &= \sum_{i=1}^{S} \Pr[\Pi = l_i] \log_2 \left(\frac{1}{\Pr[\Pi = l_i]}\right) \\ &\geq \frac{S}{n + 1} \log_2(n + 1) - 2 \geq \left(1 - \frac{\varepsilon}{\delta}\right) \log_2(n + 1) - 2. \end{aligned}$$

□

The next theorem states that for any fixed δ every two-way deterministic protocol with error probability ε must communicate about $H(X|Y) + (1 - \delta)\log_2(1/\varepsilon)$ bits on average.

Theorem 6.1. *Assume that π is a deterministic protocol which transmits X from Alice o Bob who knows Y and*

$$\Pr[\pi(X, Y) \neq X)] \leq \varepsilon.$$

Then the expected length of π is at least

$$(1 - \delta - \delta/n) \log_2 \left(\frac{\delta}{\varepsilon + \delta/n}\right) + (\delta - 2\varepsilon) \log_2(n + 1) - 2\delta.$$

The lower bound in this theorem is quite complicated and comes from its proof. To understand this bound assume that δ is a constant, say $\delta = 1/4$, and $\frac{1}{n} \leq \varepsilon \leq \frac{1}{\log_2 n}$. Then $H(X|Y) = (1/4)\log_2 n + O(1)$ and the lower bound becomes

$$\left(1 - \frac{1}{4} - \frac{1}{4n}\right)\log_2\left(\frac{\frac{1}{4}}{\varepsilon + \frac{1}{4n}}\right) + (1/4 - 2\varepsilon)\log_2(n+1) - \frac{1}{2}$$

Condition $\frac{1}{n} \leq \varepsilon$ implies that the first term is equal to

$$(3/4)\log_2\left(\frac{1}{\varepsilon}\right) - O(1).$$

Condition $\varepsilon \leq \frac{1}{\log_2 n}$ implies that the seconds term is equal to

$$(1/4)\log_2 n - O(1).$$

Therefore under these conditions the lower bound becomes

$$(1/4)\log_2 n + (3/4)\log_2\left(\frac{1}{\varepsilon}\right) - O(1) = H(X|Y) + (3/4)\log_2\left(\frac{1}{\varepsilon}\right) - O(1).$$

Proof. Let $\Pi = \Pi(X, Y)$ denote the leaf Alice and Bob reach in the protocol π for input pair (X, Y). As we have seen, the expected length of communication is at least the entropy $H(\Pi(X, Y))$. Let $l_1, \ldots, l_S$ denote all the leaves in the support of the random variable $\Pi(X, Y)$. The set $\{(x, y) \mid \Pi(x, y) = l_i\}$ is a combinatorial rectangle $R_i \subset \{0, 1, \ldots, n\} \times \{0, 1, \ldots, n\}$. Imagine $\{0, 1, \ldots, n\} \times \{0, 1, \ldots, n\}$ as a table in which Alice owns columns and Bob owns rows. Let h_i be the height of R_i and w_i be the width of R_i. Let d_i stand for the number of diagonal elements in R_i (pairs of the form (j, j)). By definition of (X, Y) we have

$$\Pr[\Pi(X, Y) = l_i] = \frac{(1-\delta)d_i}{n+1} + \frac{\delta(h_i w_i - d_i)}{n(n+1)}. \tag{4}$$

The numbers $\{\Pr[\Pi(X, Y) = l_i]\}_{i=1}^{S}$ define a probability distribution over the set $\{1, 2, \ldots, S\}$ and its entropy equals $H([\Pi(X, Y))$. Equation (4) represents this distribution as a weighted sum of the following distributions: $\left\{\frac{d_i}{n+1}\right\}_{i=1}^{S}$ and $\left\{\frac{h_i w_i}{(n+1)^2}\right\}_{i=1}^{S}$. That is, Equation (4) implies that

$$\{\Pr[\Pi = l_i]\}_{i=1}^{S} = (1 - \delta - \delta/n)\left\{\frac{d_i}{n+1}\right\}_{i=1}^{S} + (\delta + \delta/n)\left\{\frac{h_i w_i}{(n+1)^2}\right\}_{i=1}^{S}.$$

Since entropy is concave, we have

$$\begin{aligned} H(\Pi) &= H\left(\{\Pr[\Pi = l_i]\}_{i=1}^{S}\right) \\ &\geq (1 - \delta - \delta/n)H\left(\left\{\frac{d_i}{n+1}\right\}_{i=1}^{S}\right) + (\delta + \delta/n)H\left(\left\{\frac{h_i w_i}{(n+1)^2}\right\}_{i=1}^{S}\right) \end{aligned} \tag{5}$$

The lower bound of the theorem follows from lower bounds of the entropies of these distributions.

A lower bound for $H\left(\left\{\frac{d_i}{n+1}\right\}_{i=1}^{S}\right)$. In each row of R_i there is at most 1 element (x,y), for which $\pi(x,y)=x$. The rectangle R_i consists of d_i diagonal elements and hence there are at least $d_i^2-d_i$ elements (x,y) in R_i for which $\pi(x,y)\neq x$. Summing over all i we get

$$\varepsilon \geq \sum_{i=1}^{S} \frac{\delta(d_i^2-d_i)}{n(n+1)}$$

and thus

$$\sum_{i=1}^{S}\left(\frac{d_i}{n+1}\right)^2 \leq \frac{\varepsilon+\delta/n}{\delta}.$$

Since Renyi entropy is a lower bound for the Shannon entropy, we have

$$H\left(\left\{\frac{d_i}{n+1}\right\}_{i=1}^{S}\right) \geq \log_2\left(\frac{1}{\sum_{i=1}^{S}\left(\frac{d_i}{n+1}\right)^2}\right) \geq \log_2\left(\frac{\delta}{\varepsilon+\delta/n}\right).$$

In R_i, there are at most h_i good pairs (for which π works correctly). At most d_i of them has probability $\frac{1-\delta}{n+1}$. Hence

$$\Pr[\Pi = l_i, \pi(X,Y)=X] \leq \frac{(1-\delta)d_i}{n+1} + \frac{\delta(h_i-d_i)}{n(n+1)}$$

and

$$\begin{aligned} 1-\varepsilon \leq \Pr[\pi(X,Y)=X] &= \sum_{i=1}^{S}\Pr[\Pi=l_i,\pi(X,Y)=X] \\ &\leq \sum_{i=1}^{S}\left(\frac{(1-\delta)d_i}{n+1}+\frac{\delta(h_i-d_i)}{n(n+1)}\right) = 1-\delta-\delta/n+\frac{\delta}{n(n+1)}\sum_{i=1}^{S}h_i. \end{aligned}$$

The last inequality implies that

$$\sum_{i=1}^{S} h_i \geq (1-\varepsilon/\delta)(n+1)^2.$$

A lower bound for $H\left(\left\{\frac{h_i w_i}{(n+1)^2}\right\}_{i=1}^{S}\right)$. Since $h_i \leq n+1$, we have

$$\begin{aligned} \sum_{i=1}^{S}\frac{h_i w_i}{(n+1)^2}\log_2\left(\frac{(n+1)^2}{h_i w_i}\right) &\geq \sum_{i=1}^{S}\frac{h_i w_i}{(n+1)^2}\log_2\left(\frac{(n+1)^2}{(n+1)w_i}\right) \\ &= -\log_2(n+1)+\sum_{i=1}^{S}h_i\frac{w_i}{(n+1)^2}\log_2\left(\frac{(n+1)^2}{w_i}\right). \end{aligned}$$

Obviously $\frac{w_i}{(n+1)^2} \geq \frac{1}{(n+1)^2}$. By lemma 6.2 we get

$$\sum_{i=1}^{S} h_i \frac{w_i}{(n+1)^2} \log_2 \left(\frac{(n+1)^2}{w_i} \right) \geq \left(\sum_{i=1}^{S} h_i \right) \frac{1}{(n+1)^2} \log_2 \left((n+1)^2 \right) - 2$$
$$\geq (2 - 2\varepsilon/\delta) \log_2(n+1) - 2.$$

Thus

$$H\left(\left\{ \frac{h_i w_i}{(n+1)^2} \right\}_{i=1}^{S} \right) \geq (1 - 2\varepsilon/\delta) \log_2(n+1) - 2.$$

□

References

[1] Bauer, B., Moran, S., and Yehudayoff, A. Internal compression of protocols to entropy.

[2] Braverman, M., and Garg, A. Public vs private coin in bounded-round information. In *Automata, Languages, and Programming*. Springer, 2014, pp. 502–513.

[3] Braverman, M., and Rao, A. Information equals amortized communication. In *Foundations of Computer Science (FOCS), 2011 IEEE 52nd Annual Symposium on* (2011), IEEE, pp. 748–757.

[4] Brody, J., Buhrman, H., Koucky, M., Loff, B., Speelman, F., and Vereshchagin, N. Towards a reverse newman's theorem in interactive information complexity. In *Computational Complexity (CCC), 2013 IEEE Conference on* (2013), IEEE, pp. 24–33.

[5] Kushilevitz, E., and Nisan, N. *Communication Complexity*. Cambridge University Press, 2006.

[6] Newman, I. Private vs. common random bits in communication complexity. *Information processing letters 39*, 2 (1991), 67–71.

[7] Orlitsky, A. Average-case interactive communication. *Information Theory, IEEE Transactions on 38*, 5 (1992), 1534–1547.

[8] Shannon, C. E. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review 5*, 1 (2001), 3–55.

[9] Slepian, D., and Wolf, J. K. Noiseless coding of correlated information sources. *Information Theory, IEEE Transactions on 19*, 4 (1973), 471–480.

[10] Weinstein, O. Information complexity and the quest for interactive compression. *ACM SIGACT News 46*, 2 (2015), 41–64.

[11] Yeung, R. W. *Information theory and network coding*. Springer, 2008.

A The proof of Lemma 6.1

Let $\mathcal{X} \times \mathcal{Y}$ stand for the support of (X, Y). Fix $x \in \mathcal{X}$. Consider the set of all possible leafs Alice and Bob may reach in π when $X = x$. Let l_x be the leaf of minimal depth from this set. Denote the depth of l_x by $d(l_x)$. Notice that the expected length of the protocol π is at least $\mathsf{E}_{x\sim X} d(l_x)$.

Suppose that for some $x_1, x_2 \in \mathcal{X}$, $x_1 \neq x_2$ we have $l_{x_1} = l_{x_2}$. It means that there exists $y_1, y_2 \in \mathcal{Y}$ such that when $X = x_1, Y = y_1$ and when $X = x_2, Y = y_2$ Alice and Bob reach the same leaf l_{x_1}. From the rectangle property it follows that when $X = x_1, Y = y_2$ Alice and Bob reach l_{x_1} too. Hence when $X = x_1, Y = y_2$ and when $X = x_2, Y = y_2$, Bob outputs the same answer, which is contradiction.

Thus l_x defines bijection from the set of all possible values of X to some prefix-free set of binary strings. Hence $\mathsf{E}_{x\sim X} d(l_x) \geq H(X)$.

B The proof of Lemma 6.2

The function $f(x) = x \log_2 \frac{1}{x}$ increases on $[0, e^{-1}]$ and its maximum value is $e^{-1} \log_2 e < 1$. Indeed,

$$f'(x) = \frac{1}{\ln 2}(-1 - \ln x) = \frac{\ln\left(\frac{1}{ex}\right)}{\ln 2} \geq 0$$

when $x \in [0, e^{-1}]$. Since $\sum\limits_{i=1}^{k} p_i = 1$, we have

$$\#\left\{i \in \{1, \ldots, k\} \,|\, p_i > e^{-1}\right\} < e.$$

The left hand side of this inequality is an integer hence $\#\left\{i \in \{1, \ldots, k\} \,|\, p_i > e^{-1}\right\} \leq 2$. Thus we conclude

$$\begin{aligned}
\sum_{i=1}^{k} p_i \log_2 \frac{1}{p_i} &= \sum_{p_i \leq e^{-1}} p_i \log_2 \frac{1}{p_i} + \sum_{p_i > e^{-1}} p_i \log_2 \frac{1}{p_i} \\
&\geq \sum_{p_i \leq e^{-1}} q_i \log_2 \frac{1}{q_i} + \sum_{p_i > e^{-1}} 0 \\
&\geq \sum_{p_i \leq e^{-1}} q_i \log_2 \frac{1}{q_i} + \sum_{p_i > e^{-1}} \left(q_i \log_2 \frac{1}{q_i} - 1\right) \geq \sum_{i=1}^{k} q_i \log_2 \frac{1}{q_i} - 2.
\end{aligned}$$

C Random variables, for which Theorem 3.1 may be tight

We finish this paper with the example of random variables (X, Y), for which we believe that the upper bound from Theorem 3.1 is tight. Let H_n be the n-th

harmonic number:

$$H_n = \sum_{k=1}^{n} \frac{1}{k} = \ln n + O(1).$$

Let X take values in $\{1, 2, \ldots, n\}$ and Y take values in S_n, the set of all permutations of the set $\{1, \ldots, n\}$. The distribution of X, Y is defined as follows:

$$\Pr[X = i, Y = \sigma] = \frac{1}{\sigma(i) H_n n!}.$$

This formula implies that $H(X|Y = \sigma)$ does not depend on $\sigma \in S_n$ and equals

$$\sum_{i=1}^{n} \frac{\log_2(iH_n)}{iH_n} = \frac{\log_2 n}{2} + O(\log \log n).$$

Thus $H(X|Y) = \frac{\log_2 n}{2} + O(\log \log n)$.

We conjecture that every deterministic protocol, which transmits X from Alice to Bob who knows Y with error probability $\varepsilon < 1/\log_2 n$, communicates at least $\frac{\log_2 n}{2} + \Omega(\sqrt{\log_2(n)})$ bits on average.